REPORT NO. NAWCADWAR-TN-93043-60

(2)

A CRITICAL ANALYSIS OF WORKLOAD PREDICTIONS GENERATED BY MULTIPLE RESOURCE THEORY DURING EARLY CREWSTATION DESIGN

David Cohen
Air Vehicle and Crew Systems Technology Department (Code 6021)
NAVAL AIR WARFARE CENTER
AIRCRAFT DIVISION WARMINSTER
P. O. Box 5152
Warminster, PA 18974-0591

Robert J. Wherry, Jr.
The Robert J. Wherry, Jr. Co., Chalfont, PA

Floyd Glenn
Chi Systems Inc.
Spring House, PA

1 JUNE 1993

FINAL REPORT

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

PREPARED FOR
AIR VEHICLE AND CREW SYSTEM TECHNOLOGY DEPARTMENT (CODE 6021)
NAVAL AIR WARFARE CENTER
AIRCRAFT DIVISION WARMINSTER
P.O. Box 5152
Warminster, PA 18974-0591

Best
Available
Copy

NOTICES

REPORT NUMBERING SYSTEM — The numbering of technical project reports issued by the Naval Air Warfare Center, Aircraft Division, Warminster is arranged for specific identification purposes. Each number consists of the Center acronym, the calendar year in which the number was assigned, the sequence number of the report within the specific calendar year, and the official 2-digit correspondence code of the Functional Department responsible for the report. For example: Report No. NAWCADWAR-92001-60 indicates the first Center report for the year 1992 and prepared by the Air Vehicle and Crew Systems Technology Department. The numerical codes are as follows:

CODE	OFFICE OR DEPARTMENT
00	Commanding Officer, NAWCADWAR
01	Technical Director, NAWCADWAR
05	Computer Department
10	AntiSubmarine Warfare Systems Department
20	Tactical Air Systems Department
30	Warfare Systems Analysis Department
50	Mission Avionics Technology Department
60	Air Vehicle & Crew Systems Technology Department
70	Systems & Software Technology Department
80	Engineering Support Group
90	Test & Evaluation Group

PRODUCT ENDORSEMENT — The discussion or instructions concerning commercial products herein do not constitute an endorsement by the Government nor do they convey or imply the license or right to use such products.

Reviewed By: J. K. Schmidt ________ Date: 20 OCT 93
Branch Head LCDR MSC USN

Reviewed By: ________ Date: 20 Oct 93
Division Head

Reviewed By: Michael J. ________ Date: 21 Oct. 93
Director/Deputy Director

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	1 June 1993	Final

4. TITLE AND SUBTITLE

A CRITICAL ANALYSIS OF WORKLOAD PREDICTIONS GENERATED BY MULTIPLE RESOURCE THEORY DURING EARLY CREWSTATION DESIGN

5. FUNDING NUMBERS

6. AUTHOR(S)

David Cohen*
Robert J. Wherry, Jr.**
Floyd Glenn***

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

* Air Vehicle and Crew Systems Technology Department (Code 6021)
NAVAL AIR WARFARE CENTER
AIRCRAFT DIVISION WARMINSTER
P. O. Box 5152
Warminster, PA 18974-0591

8. PERFORMING ORGANIZATION REPORT NUMBER

NAWCADWAR-93043-60

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Air Vehicle and Crew Systems Technology Department (Code 6021)
NAVAL AIR WARFARE CENTER
AIRCRAFT DIVISION WARMINSTER
P. O. Box 5152
Warminster, PA 18974-0591

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

Robert J. Wherry, Jr. **	Floyd Glenn ***
The Robert J. Wherry, Jr. Co.	Chi Systems Inc.
Chalfont, PA	Spring House, PA

12a. DISTRIBUTION / AVAILABILITY STATEMENT

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

Subjective workload ratings based on multiple resource theory were independently collected from two highly experienced pilots for 225 different tasks of an anticipated mission for a future advanced strike aircraft. Factor analysis of their responses suggest that while such ratings could have little actual validity in and even high inter-rater reliabilities, the ratings have high face validity and even high inter-rater reliabilities, the ratings could have little actual validity in terms of efforts required to utilize the seven postulated resource channels (visual or auditory input, spatial, verbal, or analytical cognition, and manual or speech output). Ratings of efforts required for various postulated cognitive resource channels were particularly suspect. Four independent factors were identified for each pilot which accounted for virtually all of the intercorrelations among the seven resource channels. Three factors (visual-spatial, verbal communications, and manual and speech output) were identical for both pilots and accounted for most of their explainable variance.

14. SUBJECT TERMS	15. NUMBER OF PAGES
Lighting, AComposite, Conductive Paint, Electrical Bonding Graphite, V-22, SCMCM	65
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	SAR

ABSTRACT

Subjective workload ratings based on multiple resource theory were independently collected from two highly experienced pilots for 225 different tasks of an anticipated mission for a future advanced strike aircraft. Factor analysis of their responses suggest that while such ratings have high face validity and even high inter-rater reliabilities, the ratings could have little actual validity in terms of efforts required to utilize the seven postulated resource channels (visual or auditory input, spatial, verbal, or analytical cognition, and manual or speech output). Ratings of efforts required for various postulated cognitive resource channels were particularly suspect. Four independent factors were identified for each pilot which accounted for virtually all of the intercorrelations among the seven resource channels. Three factors (visual-spatial, verbal communications, and manual and speech output) were identical for both pilots and accounted for most of their explainable variance.

In an effort to increase efficiency and lower costs, military design programs are increasingly emphasizing upfront analyses, including predictions of operator workload. It is critical that reasonable forecasts of operator performance be made prior to full scale development so as to avoid costly delays and subsequent design changes. Unfortunately, this situation mandates that these analyses be conducted prior to the establishment of a concrete baseline design with measurable human performance variables. Therefore, to provide operator workload assessments early in the design process, most methods simply require a sample of prospective operators to project themselves into the future system and rate the amount of physical and mental demands they expect during system employment. Often, subjective estimates are considered within the context of a model of human performance to produce more "realistic" and systematic projections of task effects.

These models partition high-level human functionality (i.e., perception, cognition and motor action) into lower-level dimensions which are more readily translated into design decisions. For example, perception may be broken into vision, audition and touch - each of which can be related differently to control and display solutions.

Although many possible problems have been identified in connection with using subjective opinion of workload (for a review see Williges & Wierwille, 1979), one has been overlooked. Most operators believe they can easily rate their predicted capabilities, and their ability to discriminate between their different capabilities does have face validity. However, it is crucial to know if they are *actually* discriminating the various human resources according to the differential impacts of task requirements if subjective methods are to have good predictive validity. This paper describes an experimental subjective workload analysis undertaken at the Naval Air Warfare Center and a subsequent critical examination of the results to determine what was *actually* being rated.

BACKGROUND

In an effort to predict pilot workload early in the crew station design process, the Advanced Technology Cockpit (ATC) Pilot-Vehicle Interface (PVI) program incorporated a workload measure based on the Workload Index (W/INDEX) model (North & Riley, 1988) into a task network simulation of an advanced strike mission (Hodorovich & Cohen, 1989). The W/INDEX model is predicated on multiple resource theory (Wickens, 1984). This theory states that humans possess several distinct resources or channels (to perform tasks) rather than one undifferentiated pool of resources. Total workload at any time will be the sum of the loading on each of the distinct channels plus any penalties incurred across channels (conflicts). Time sharing of resources will occur to the extent that simultaneously occurring tasks place demands on different resources. The W/INDEX algorithm calculates workload across human resources (e.g., vision, audition, etc.), considering between resource (e.g., visual by auditory) and within resource (e.g., visual by visual) conflicts as well as additive workload given subjective estimates of the impacts a task has on these resources.

These impacts must be considered across an accurate representation of the pilot's activities in the cockpit. Our approach used task network modeling to construct simulations of the strike mission. A task network model differentiates human performance into a series of subtasks with the relationships between subtasks defined by a network which connects them (Chubb, Laughery, & Pritsker, 1987). In more elementary terms, a task network is a hierarchical grouping of subtasks. The structure of the task network specifies the order of execution of subtasks as well as their branching to subsequent subtasks. Mathematical or logical expressions, like the W/INDEX algorithm, can be embedded in the simulation and thus operate on the values (i.e., the resource estimates) that are active through the proper paths at the proper times in the simulation.

Most design programs are satisfied with the outputs of such a simulation (i.e., relative workload values across time); however, we proceeded to critically analyze the products of this simulation. Factor analyses and Multiple Correlation studies of the results revealed that people may be limited in their ability to discriminate between discrete influences on their task performance.

The remainder of this article will briefly recount the methodology that we employed to implement a W/INDEX-like model into a task network simulation, the results of that simulation, the analysis of those results and possible steps that can be taken to address the problems that were encountered.

METHOD

Subject Matter Experts

Resource effort estimates were provided by two recently retired U.S. Marine Corps pilots (P1 and P2 individually). Both of these pilots had significant operational experience (approximately 1000 hours) in the F/A-18 Hornet, which is an antecedent to the next-generation fighter/attack aircraft, as well as combat experience in the F-4 Phantom II. In addition, both pilots had assisted in the development of the strike mission scenario and the stipulation of the aircraft capabilities and therefore were intimately familiar with the tasks that were rated.

Workload Estimation

The pilots were asked to rate the amount of effort that will be required in each of seven human resources or channels in order to perform each of 225 strike tasks. These channels included: visual perception, auditory perception, spatial information processing, analytical information processing, verbal information processing, manual activity, and speech. An eight point scale was used in which "0" indicated "no effort required" and "7" indicated "maximum effort required." They were also requested to estimate the overall effort needed to complete the task without the partitioning of resources. The pilots were instructed to rate each task and/or each component of a task independent of any concurrent task or component. These estimates were gathered and recorded using a HyperCard program running on a Macintosh SE computer. Figure 1 shows the display interface used for data collection. Details on the definition of the resource categories, the data collection procedures and the construction of the data collection system can be found in Glenn, Cohen, Barba, and Santerelli (1990).

Insert Figure 1 about here

Network Simulation Construction

MicroSaint simulation software running on a 386 personal computer was used to implement task network representations of the strike mission. MicroSaint, a product of Micro Analysis and Design Inc., allows the user to develop, execute, and analyze the results of network simulation models. Models are constructed by defining task nodes and connecting them together via branching or control logic to form a task network. A task node consists of its associated attributes, which usually includes: task identification, mean execution time, beginning and ending effects, and following task information. When the simulation is executed, the software provides the ability to capture data on the state of the simulation. For a more comprehensive description of MicroSaint and its application to a tactical mission (for the LHX helicopter) see Laughery, Drews, and Archer (1986).

The required models were constructed for each of the ten _phases_ of the strike mission: take-off, climb, cruise out, descent, ingress, attack, egress, climb (second), return to force, and recovery. The timeline for each phase was further decomposed into _segments_ within mission phases (e.g., aviate, navigate, etc.) and individual _tasks_ (e.g., monitor system status) using the original ATCS task analyses as a reference (Cohen, 1990). The models were developed from an analysis of the ATCS strike mission timelines (Veda, 1990). Task networks were then created by assigning connections between tasks on the basis of task execution times and logical heuristics. Task start times and durations were acquired from the timelines and later verified by subject matter experts. Mission segments were used as the starting point for all tasks within that segment. In the models, mission segments can be considered pseudo-tasks because although they have no time or workload charges associated with them, they were needed to provide the grouping for tasks. Figure 2 shows an example of the network diagrams that were drawn to represent the structure of the task relationships (see Glenn et al., 1990).

Insert Figure 2 about here

After the task networks diagrams were developed, they were implemented in MicroSaint. Network models were built using the task connections shown in the network diagrams and the task timing information obtained from the timelines. The release condition for each task contains a function (i.e., logical and mathematical control statement) which forces the task to execute at the correct time to effectively mimic the timeline. Mean execution times for tasks were taken directly from the timelines. When tasks repeated more than once with different task durations, a variable was inserted as the mean time. Functions were written to insert the correct time value into the mean time variable at the appropriate time. Task beginning effects contained the workload values across the seven channels (described below) for all the tasks. When a task was executed, its associated workload values became active which caused them to be included in the workload calculation. Task ending effects contained zeros for all channels to initialize the task workload values. Tasks which could follow execution of some other task were assigned on the basis of the examination of the timelines. The probability of taking a following task contained functions which controlled branching to other tasks or back to itself, if that task was iterative.

The simulations were set to use a one second time step so that workload would be calculated for each second. In addition to workload (which is defined as the total loading according to the W/INDEX equation), individual channel loading values were also captured at one second intervals. The simulations which were created in this effort were both fully deterministic and clock-driven. The simulations will yield the same results each time they are run and these results are tied directly to the clock. This was done to ensure that all tasks begin and end at the correct time and conform to the ATCS strike timeline.

Workload Model

The function to calculate workload based on the subjective ratings was the instantiation of the W/INDEX algorithm. Total workload was divided into components based on the SMEs' estimates of the effort taxing the seven resources. The first two channels (visual and auditory) represented input channels. The next three channels (spatial, analytical, and verbal) represented cognitive processing channels. The last two channels (manual and speech) corresponded to output channels. Within each task network, all tasks were assigned workload values for each of the seven channels. These values were valid for the duration of the task.

The W/INDEX algorithm used these estimates to calculate workload according to the following expression:

$$W_T = \sum_{i=1}^{l} \sum_{t=1}^{m} a_{t,i} + \sum_{i=1}^{l} \left[(n_{t,i} - 1) C_{ii} \sum_{t=1}^{m} a_{t,i} \right] + \sum_{i=1}^{l-1} \sum_{j=i+1}^{l} C_{ij} \sum_{t=1}^{m} (a_{t,i} + a_{t,j})$$

where:

W_T	=	instantaneous workload at time T
i, j	=	1...l are the resource channels
t	=	1...m are the tasks occurring at time T
$n_{t,i}$	=	number of tasks occurring at time t with nonzero load values for channel i
$a_{t,i}$	=	load value for channel i in performing task t
$a_{t,j}$	=	load value for channel j in performing task t
$c_{i,j}$	=	conflict between channels i and j
c_{ii}	=	conflict within channel i

(NOTE: The third term of the W/INDEX algorithm is only calculated when both $a_{t,i}$ and $a_{t,j}$ are non-zero.)

One of the major features of the W/INDEX algorithm is its use of a conflict matrix to assess the workload penalties associated with the concurrent activity of any two channels or the use of a single channel by concurrent tasks. The conflict matrix that was used in these simulations consisted of 28 terms which represent the conflict of each of the seven channels with itself and all other channels. The conflict coefficients (figure 3) were adapted from the research of North and Riley (1988) and ranged from 0 to 1. A technical discussion of the implementation of the features of multiple resource theory into the task network simulation (including the function source code) can be found in Glenn et al. (1990).

Insert Figure 3 about here

RESULTS

Workload Predictions

The purpose of this article is to present the analysis of the workload predictions generated by multiple resource theory as opposed to the predictions themselves (see Glenn et al., 1990 for the complete workload predictions). Sample outputs of the simulation are included in figures 4, 5, and 6. These figures show the diversity of the outputs that were available in the implementation, including: total instantaneous workload (figure 4), individual channel loadings (figure 5) and the contributions of the conflict matrix (figure 6). It is important to note the extreme range and non-linearity of the workload predictions.

Insert Figures 4, 5, & 6 about here

Correlations and Factor Analysis

Means, standard deviations, and correlations of the workload ratings of the seven resources across all tasks were obtained independently for both P1 and P2. Relatively high intercorrelations among all seven resource channels and extremely high correlations among some of them suggested that raters must have felt that many tasks required all of the "independent" resource channels or that the raters were unable to discriminate among them. At the very least, the raters appeared to be indicating that whenever high effort levels were required by any input resource channel, high effort levels would also be required for cognitive and output channels as well. To identify the number and nature of independent factors causing the high intercorrelations among the seven postulated resource channels, Principal-Axis (PA) factor analyses of the intercorrelations for each subject were accomplished. For these analyses, initial communalities (h^2s) for each factor analysis were estimated using the highest-r method. Solutions were iterated until beginning and ending communality estimates stabilized within .001. Four factors were extracted for each pilot. Varimax-rotated factors failed to yield simple structure (i.e., where some variables have high loadings on a factor and all others have zero loadings) for the factors for either pilot. Ultimately, graphical rotation was used to identify the general factor responsible for the extremely high intercorrelations among the seven resource channels. Results of those analyses are shown in Table 1.

Insert Table 1 about here

The sum of the eigenvalues (i.e., the sum of the resource channels' variance explained by each factor) and the sum of the communalities (i.e., the sum of each variable's variance explained by all of the factors) show that 92.6% (i.e., = 6.481/7) of the variance of all variables across all tasks was explained by P1's four factors. For P2, the comparable figure was 73.4% (i.e., = 5.140/7).

<u>Interpretation of the Rotated Factors</u>

Both pilots yielded a very strong general factor (i.e., one in which all variables have high loadings) that loaded most highly (.973 and .982, respectively) with the visual input channel. The second highest loadings on those factors was the spatial information processing channel (.981 and .787). This indicates that both pilots perceived that when the tasks being rated were dominated by visual inputs, they also required spatial processing.

Because all of the other channels loaded significantly on this **visual-spatial factor** (factor 1), it indicates that the tasks dominated by visual-spatial demands were sufficiently complex to demand the other resource channels as well (e.g., analytical thought, verbal communications, and manual outputs).

A second and independent **verbal-communications factor** (factor 2) was also found for both pilots. It was dominated by high loadings on auditory input, verbal information processing, and speech output. This factor indicates that the pilots also distinguished tasks that were dominated by (or required relatively more or less) verbal communications.

A third and independent **manual and speech output factor** (factor 3) was also found for both pilots, although with somewhat weaker loadings for P1. This factor indicates that the pilots distinguished among tasks that required relatively more or less output demands.

While an additional independent factor was found for each pil : (factor 4), the nature of their final factors appeared to be quite different. For P1, the final factor loaded highest on verbal information processing (.349) and speech output (.438) indicating P1 differentiated among tasks that required more or less **speech production** than would have been indicated by the loadings for the resources on the visual-spatial or verbal-communications factors. For P2, the final factor had high loadings on the analytical (.620) and spatial (.395) information processing channels, indicating that P2 may have made finer distinctions concerning the amount of **analytical thought required for spatial tasks**.

By far the most variance of the ratings for both pilots was explained by the first factor. This suggests that differential workload ratings (at least for these tasks) were determined primarily on the basis of the extent to which the visual-spatial factor was important to the rated tasks.

Further Analysis of the Factors

To determine the relative importance of each factor to the tasks, the seven channels for each pilot were used as predictor variables for each of the four factors using a multiple correlation program. The resulting prediction equations (assuming standard scores are desired for each factor) are shown in Table 2.

Insert Table 2 about here

Using the prediction equations shown in Table 2, the factor scores for each pilot were then computed. For each pilot, the correlations of the four factor scores for each task along with the ratings of the seven channels were computed. The factors were then used as predictors for each channel variable. The resulting prediction equations are shown in Table 3. These equations in conjunction with the previously calculated factor scores were then used to compute estimates of each of the resource ratings for each task. As would be expected from the multiple Rs reported in Table 3, the predicted ratings for each channel for all tasks were very close to the actual ratings for both pilots. For P1, over 91% of the predicted ratings were within .5 of the actual ratings while nearly 98% were within 1.0 of the actual ratings. For P2, over 75% of the predicted ratings were within .5 of the actual rating, and over 92% were within 1.0 of the actual ratings. Thus, the predicted ratings, based only on four dimensions, closely predicted the ratings given by each subject on the seven postulated resource channels' seven-point rating scales.

CONCLUSIONS

While the seven postulated resource channels may represent independent capabilities, it is clear that their rated usages were highly related for the 225 tasks studied. Further, visual-spatial, verbal communications, and output factors (which together accounted for a very large proportion of the variance of the ratings as well as most of the correlations among the resource channels) emerged for both Ss. This strongly suggests that the seven channels, even if they do represent independent resources, are strongly confounded in real-world tasks. For example, it is not surprising to expect that much of the task information presented visually would require some sort of spatial processing.

Nor is it surprising to find that many speech communications tasks would involve auditory inputs, internal verbal processing, and speech outputs. Finally, it is not surprising to find that some tasks may require differential amounts of information outputting relative to the amount of information input and processed internally. For example, monitoring and supervisory tasks usually require only occasional information outputs relative to inputs.

More problematic is the issue of the extent to which these stereotypes of required combinations of resources are being imposed by the raters on the tasks. For example, raters may be assuming that tasks requiring a certain level of workload for inputting visually presented information must also require similar amounts of workload for internal spatial processing, or that tasks requiring certain levels of speech inputs and outputs must also require similar amounts of internal verbal processing. If this is the case, then ratings of workload for the internal (cognitive) processes in this study may simply be the result of beliefs rather than independent assessments of time or effort to accomplish those internal processes. Such stereotypes could easily account for the high correlations found in this study.

DISCUSSION

The results from this study strongly suggest that although subjective opinions of projected workload may have high face validity, especially when collected from subject matter experts, these estimates may not be valid indicators of the real effort levels that will be required of operators when the actual system has been developed. We have no particular quarrel with the general concept of multiple resources (i.e., input, cognitive, and output) being required for the accomplishment of most real-world tasks. Further, we believe that it is relatively easy for subject matter experts to determine the types of inputs (visual or auditory) and outputs (manual or speech) required for any task. Identifying and distinguishing among the types of cognitive resources needed for a task may be somewhat more difficult. However, in eliciting opinions about projected workload, we are not merely asking the rater to identify the types of resources needed, but to tell us the level of effort needed for each resource. The high positive loadings for all seven resource channels on factor 1 for both pilots suggest that our subject matter experts may simply have arrived at an overall estimate of how difficult they thought a particular task would be and then justified that belief by assigning what seemed to them to be appropriately high or low effort ratings to all of the resource channels. In rating parlance this would be referred to as a halo effect. Our findings also indicate that the subject matter experts can make distinctions between the types of inputs and outputs required of the task. This is evidenced by the fact that for both pilots, factor 1 was dominated by extremely high loadings for visual inputs and manual responses while factor 2 was dominated by auditory inputs and speech responses. The estimates of effort required for various cognitive resource channels given by raters appear to be based almost solely on the type of inputs they thought were important to the task. Thus, postulated cognitive resource channels appear to have received effort ratings proportional to the input channels the raters believed to be related to those channels.

The above discussion suggests that workload estimation methods such as W/INDEX, when based on a very limited number of input, cognitive, and output resources, and when used as a prospective workload technique, may generate data that have high face validity (and even high reliability and general agreement across raters). However, the ratings may not actually provide valid indications of the actual workload efforts that will be required when the system has finally been developed.

The complexity of the W/INDEX equations (its workload model) and its utilization of conflict matrices certainly give it the appearance of a carefully constructed and precise instrument for determining workload. When the W/INDEX model is further coupled with a task simulation network program, together they can produce a variety of apparently sophisticated outputs (e.g., total instantaneous workload, individual channel loadings, etc.) which, while costly to achieve, may not provide the diagnostic utility they purport to yield. Before these types of prospective workload estimation techniques become widely adopted, we need studies demonstrating that early projected estimates of efforts required for system tasks do, in fact, correlate highly with actual efforts required by those same tasks. This study did not attempt to do this since the system we studied has yet to be developed.

One of the stated reasons we attempt to obtain early workload projections is to determine whether operators will have sufficient resources, in terms of capabilities, effort, and time, to accomplish all of their allocated tasks. In our study, the task time needed (or available) to perform each task had been pre-estimated (as part of the mission timeline) independently of effort required for that task. Our emphasis in this study on determining the effort levels required for those same tasks follows the contention by Stewart and Lofaro (1990) that a key determinant of workload is effort required, or the difficulty of the task and how long it must be performed since both tie up resources. However, they report that while a high correlation ($\underline{r}$ = .93) has been found by Gopher and Braune (1984) between workload estimates and subjective ratings of task difficulty, the correlation between workload estimates and actual performance times was fairly low ($\underline{r}$ =.30). Thus, our prospective workload estimates were based solely on efforts required to perform various types of activities for each task within a stated amount of time.

If the separate resource effort ratings also turn out to have low correlations with actual times required to perform those same activities, then that would limit the value of the ratings for identifying and resolving conflicts where independent activities compete for the same resources during a given time period.

In adopting the multiple resource theory as part of the W/INDEX model, the workload rater is asked to go beyond describing overall effort required by a task and, instead, describe the effort levels required for a variety of different perceptual, cognitive, and response activities. Our data suggest that raters, when evaluating systems that have yet to be developed, are limited in their abilities to distinguish separate performance resources that might be required, especially in the cognitive domain. Further studies would also be useful to determine the extent of correlation between both projected times and effort levels and (once the system is developed) the actual times and subjective effort levels expended for each of the resource channels.

We recognize that the concept of workload is broader than the concept of performance time and accuracy. With workload we desire to know how close we are coming to overloading the capacity of the operator rather than simply if the operator will be able to perform all of the assigned tasks. If multiple resource approaches are to be taken with regard to estimating overall task effort and in discriminating among different types of activities which lead to operator overload, then it would seem equally reasonable to first enquire as to the percentages of overall allocated task times that must be dedicated to each activity type. Elicitation of these types of responses should more directly identify multi-channel, multi-task resource conflicts. However, as suggested earlier, before we adopt such techniques, we do need data to demonstrate that these kinds of projected estimates can be validly made by raters.

REFERENCES

Chubb, G., Laughery, R., & Pritsker, A. (1987). Simulating manned systems. In G. Salvendy (Ed.), Handbook of human factors (pp. 1298-1327). New York, NY: John Wiley & Sons.

Glenn, F., Cohen, D., Barba, C., & Santerelli, T. (1990). The advanced technology crew station: Initial workload assessment (NADC Technical Report No. NADC-90117-60). Warminster, PA: Naval Air Development Center.

Gopher, D. & Braune, R. (1984). On the psychophysics of workload: why bother with subjective measures? Human Factors, 26, (pp. 519-532).

North, R. & Riley, V. (1989). W/INDEX: A predictive model of operator workload. In G. MacMillan, D. Beevis, E. Salas, M. Strub, R. Sutton, & L. Van Breda (Eds.), Applications of human performance models to system design. (pp. 81-89). New York, NY: Plenum Press.

Stewart, J. & Lofaro, R. (1990). A secondary analysis comparing subjective workload assessments with U.S. Army aircrew training manual ratings of pilot performance. In Proceedings of the Human Factors Society 34th Annual Meeting. (pp. 104-108).

Wickens, C. (1984). Engineering psychology and human performance. Columbus, OH: Charles E. Merrill Publishing Company.

Williges, R. C. & Wierwille, W. W. (1979). Behavioral measures of aircrew mental workload. Human Factors, 21, 549-574.

Table 1

Results of Factor Analysis and Graphical Rotations

Data analysis results for pilot 1 (P1):

resource			residuals & correlations							factor loadings				
channel	Mean	S.D.	1	2	3	4	5	6	7	1	2	3	4	h^2
1 visual	3.13	2.05		585*	954	673	924	928	779	**973**	008	041	-031	949
2 auditory	1.21	1.40	-002		570	805	628	566	628	**597**	**689**	007	001	831
3 spatial	2.99	2.23	000	001		653	930	908	764	**981**	-024	-045	-036	966
4 verbal	1.26	1.39	003	000	-001		727	657	835	**693**	**568**	013	**349**	925
5 analytical	2.87	1.98	000	001	001	-002		891	811	**951**	086	-008	057	915
6 manual	2.72	2.01	002	000	-001	-001	-001		818	**940**	004	**295**	002	971
7 speech	1.36	1.76	-003	000	001	000	001	001		**807**	**210**	**193**	**438**	924
eigenvalues										5185	850	128	319	6481
portion of total variance										74.0	12.1	1.8	4.6	92.6

Data analysis results for pilot 2 (P2):

resource			residuals & correlations							factor loadings				
channel	Mean	S.D.	1	2	3	4	5	6	7	1	2	3	4	h^2
1 visual	2.90	1.63		448	788	578	515	559	276	**982**	002	027	019	965
2 auditory	1.38	1.30	-002		391	517	393	320	539	**453**	**579**	124	012	556
3 spatial	2.94	1.95	002	009		497	650	548	309	**787**	-002	177	**395**	807
4 verbal	1.97	1.41	003	000	-011		503	323	400	**582**	**434**	000	127	544
5 analytical	2.50	1.50	-002	-005	001	007		281	285	**513**	273	003	**620**	722
6 manual	2.03	1.73	-001	-005	-001	006	000		556	**553**	-010	**646**	-003	723
7 speech	.93	1.23	001	003	-001	-003	001	001		**261**	**584**	**644**	-020	824
eigenvalues										2765	939	879	557	5140
portion of total variance										39.5	12.1	12.6	04.6	73.4

* three decimals omitted for values other than means and standard deviations and variance portion

Table 2

Results of Multiple Correlation to Predict Factor Scores

Multiple Rs and factor score prediction weights for pilot 1 (P1)

factor	multiple R	resource channels 1	2	3	4	5	6	7	constant
factor 1	.9930	.119	-.005	.211	.012	.073	.064	.014	-1.416
factor 2	.9033	-.078	.505	-.192	.487	-.022	-.015	-.163	-.082
factor 3	.8291	-.168	.039	-.577	-.158	-.182	1.012	.124	.004
factor 4	.8358	-.118	-.381	-.093	.294	.027	-.301	.700	.526

Multiple Rs and factor score prediction weights for pilot 2 (P2)

factor	multiple R	resource channels 1	2	3	4	5	6	7	constant
factor 1	.9832	.571	.021	.020	.018	-.006	.011	-.032	-1.755
factor 2	.8325	-.119	.275	-.119	.195	.123	-.189	.435	-.396
factor 3	.8539	-.261	-.094	.113	-.100	-.077	.344	.421	-.143
factor 4	.8052	-.452	-.066	.316	.014	.446	-.006	-.125	-.543

Table 3

Multiple Correlations Results to Predict Ratings from Factor Scores

Multiple Rs and resource channel prediction weights for pilot 1 (P1):						
resource	multiple	factors				
channel	R	F1	F2	F3	F4	constant
visual	.9811	2.026	.003	.040	-.097	3.134
auditory	.9738	.828	1.206	.063	-.228	1.213
spatial	.9941	2.229	-.101	-.244	-.078	2.992
verbal	.9847	.966	.866	-.016	.536	1.262
analytical	.9639	1.914	.150	-.111	.141	2.867
manual	.9992	1.896	.024	.787	-.067	2.721
speech	.9900	1.424	.284	.358	1.018	1.364
Multiple Rs *and* resource channel prediction weights for pilot 2 (P2)						
resource	multiple	factors				
channel	R	F1	F2	F3	F4	constant
visual	.9993	1.655	-.002	.009	-.033	2.897
auditory	.8346	.609	1.089	.002	-.087	1.380
spatial	.9478	1.539	-.188	.500	1.145	2.940
verbal	.8012	.846	.904	-.184	.162	1.968
analytical	.9535	.750	.503	-.061	1.355	2.500
manual	.9444	.950	-.328	1.561	.028	2.024
speech	.9791	.311	.867	.916	-.084	.926

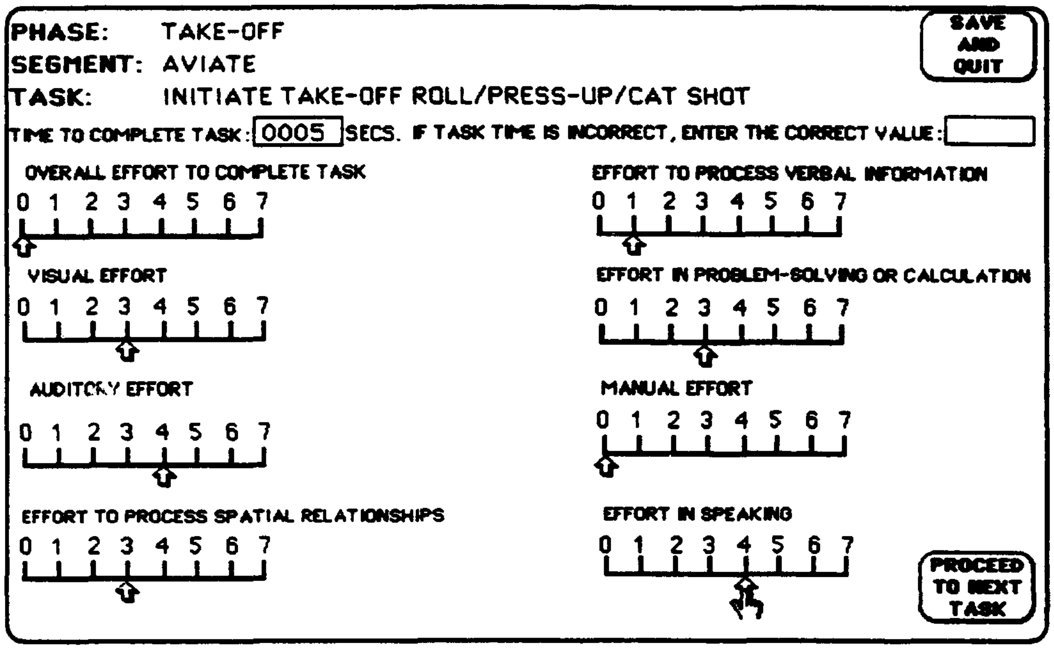

FIGURE 1. Screen Used to Elicit Resource Effort Estimates

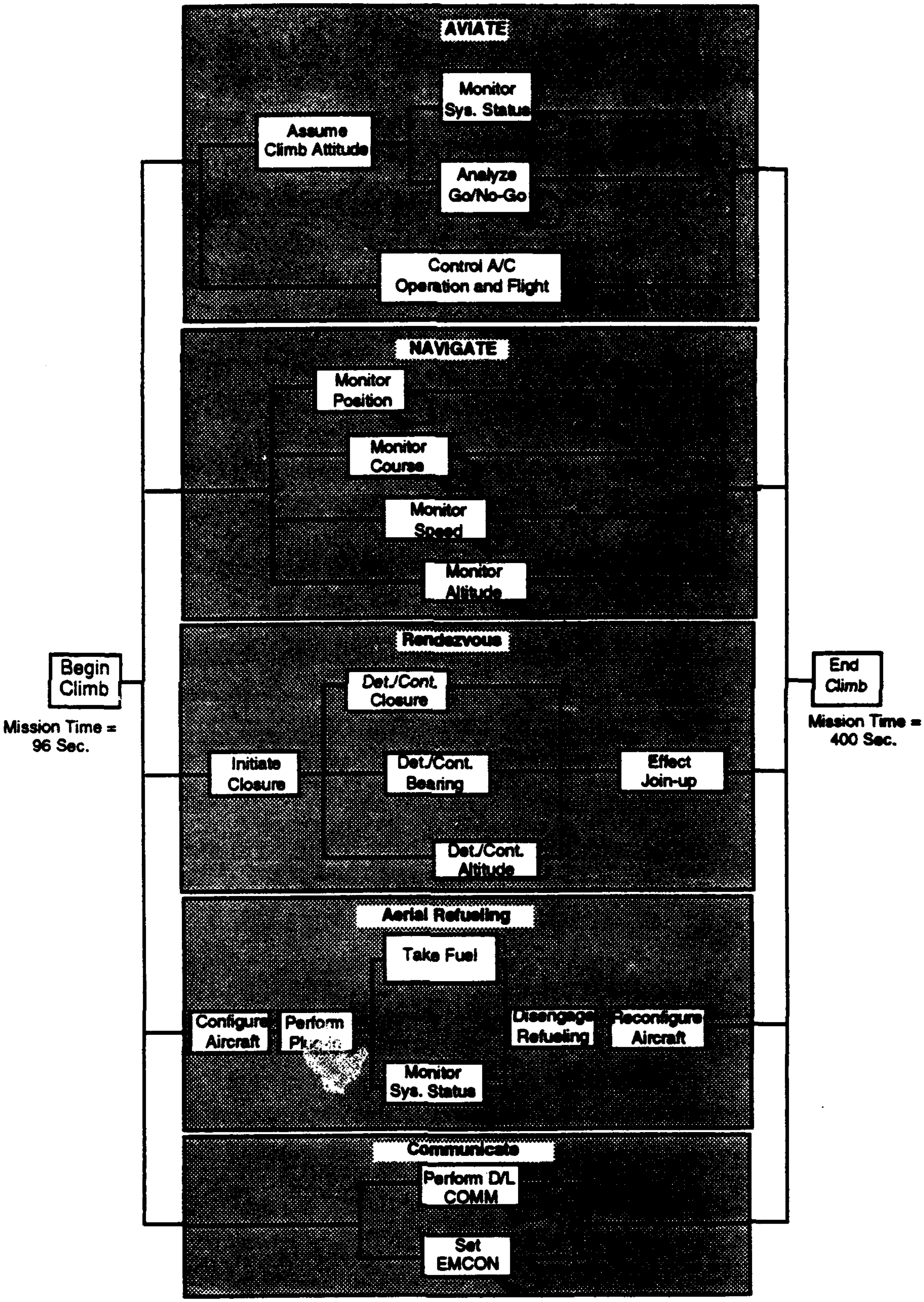

Figure 2. Climb Phase Task Network Diagram

		Input		Internal Processing			Output	
		Visual	**Auditory**	**Verbal**	**Spatial**	**Analytical**	**Manual**	**Speech**
Input	**Visual**	0.8	0.3	0.2	0.3	0.2	0.2	0.2
	Auditory		0.8	0.3	0.2	0.2	0.2	0.5
Internal Processing	**Verbal**			0.2	0.2	0.2	0.2	0.3
	Spatial				0.3	0.3	0.3	0.2
	Analytical					0.1	0.2	0.2
Output	**Manual**						0.8	0.3
	Speech							1

Figure 3. Conflict Matrix

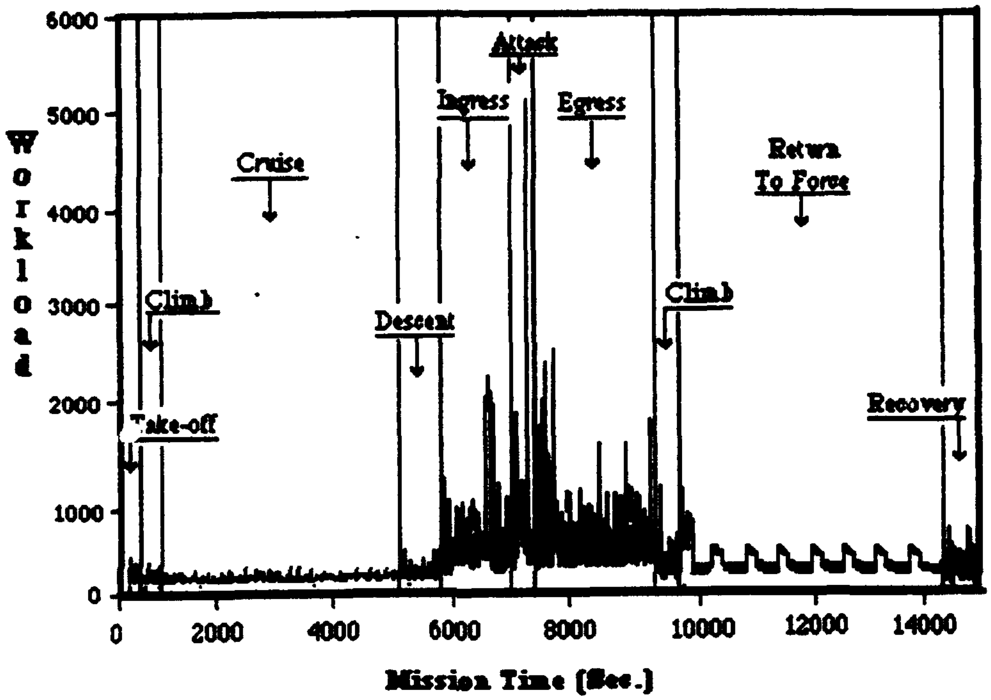

Figure 4. Instantaneous Strike Mission Workload

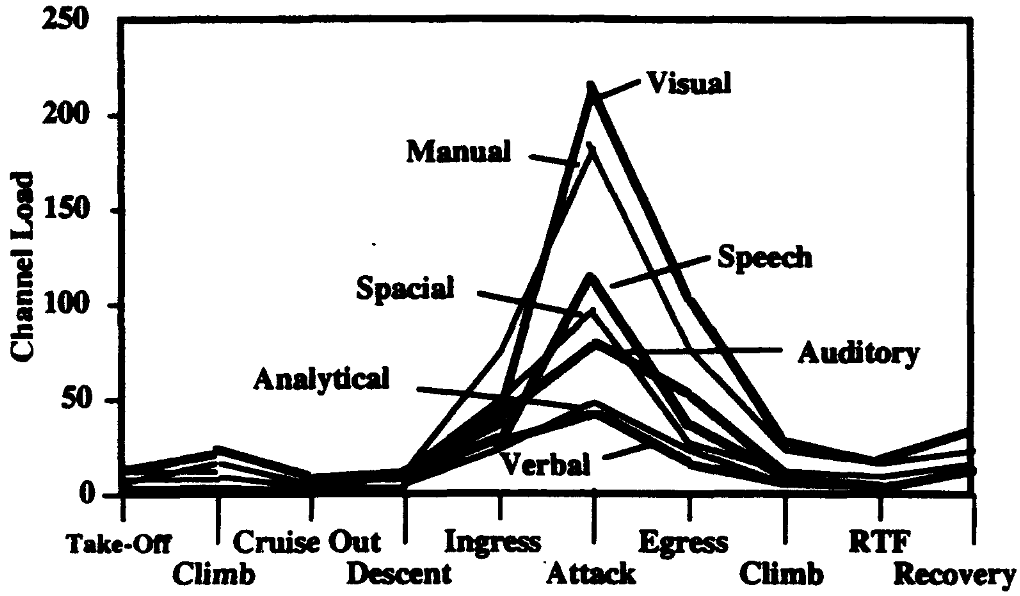

Figure 5. Average Channel Loading by Segment

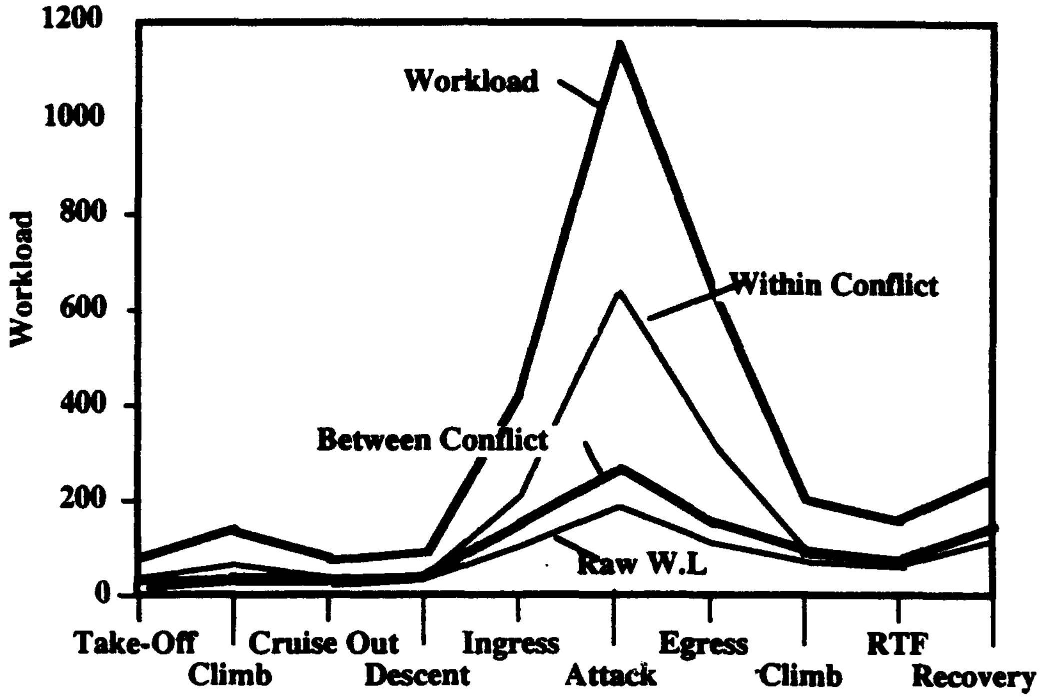

Figure 6. Average Workload and W/INDEX Components